doi:10.11959/j.issn.1000-436x.2017116

基于寻找小重量码字算法的 LDPC 码开集识别

于沛东，彭华，巩克现，陈泽亮
（解放军信息工程大学信息系统工程学院，河南 郑州 450001）

摘 要：LDPC 码的开集识别是信道编码识别领域的一个难点。首先，对实现开集识别所需接收码向量的数量进行了分析，给出了其理论下界。然后，根据这一下界，基于寻找小重量码字的算法，提出了一种新的 LDPC 码开集识别方法。该方法在接收码向量空间的对偶空间中逐个寻找小重量向量，即待识别的稀疏校验向量，从而重建稀疏校验矩阵。利用指数分布对迭代次数进行建模，给出了该方法的迭代停止准则及运算量分析。在无误码条件下，新方法克服了已有方法在适用范围和所需数据量的局限。在有误码条件下，与已有方法相比，在提高抗误码能力的同时保持较低的运算复杂度，更能满足实际应用的需求。对于 QC-LDPC 码，利用其稀疏校验矩阵的准循环特性，可以显著提高识别性能。

关键词：信道编码识别；LDPC 码；准循环 LDPC 码；指数分布
中图分类号：TN911.7 **文献标识码**：A

LDPC code reconstruction based on algorithm of finding low weight code-words

YU Pei-dong, PENG Hua, GONG Ke-xian, CHEN Ze-liang
(School of Information Systems Engineering, PLA Information Engineering University, Zhengzhou 450001, China)

Abstract: LDPC code reconstruction without a candidate set is one of the tough problems in channel code reconstruction. First, theoretical analysis was provided for the number of received code-vectors needed for the reconstruction, and a lower bound was derived. Then, according to the lower bound, and based on an algorithm for finding low weight code-words, a new reconstruction method was proposed. It looked for low weight vectors one by one from the dual space of the received code-vector space and used them to reconstruct the sparse parity-check matrices. Number of iterations and the computational complexity of the method were analyzed based on exponential distribution theory. Under noise-free conditions, drawbacks of the existing method, including limited applicable range and large quantity of required data, have been overcame. Under noisy conditions, the proposed method has higher robustness against noise and relatively low complexity, compared to existing methods. For QC-LDPC codes, the reconstruction performance can be further improved using the quasi-cyclic property of their sparse parity-check matrices.

Key words: channel code reconstruction, LDPC code, quasi-cyclic LDPC code, exponential distribution

1 引言

近年来，信道编码识别问题成为一个研究热点。信道编码识别是指根据接收解调后得到的编码序列，逆向识别出信道编码所采用的参数。这项技术在非合作通信、认知无线电等领域具有重要意义。目前，针对 BCH 码、RS 码、卷积码、Turbo 码、扰码等编码类型，都已提出了较丰富的识别方法，达到了较好的识别效果[1~5]。LDPC 码[6,7]是被各种通信标准和协议广泛采用的一类新型高效编码，具有逼近香农限的优良纠错性能。然而，对于 LDPC 码识别的研究，目前主要限于闭集识别[8,9]，对于它的开集识别的研究成果很少，且性能并不理想[10~14]。本文主要研究 LDPC 码的开集识别问题。

收稿日期：2016-08-07；**修回日期**：2017-04-18
基金项目：国家自然科学基金资助项目（No. 61401511）
Foundation Item: The National Natural Science Foundation of China (No.61401511)

闭集识别是指已知由若干 LDPC 码构成的一个集合（如已知发送方所采用的协议，其中仅规定了数种 LDPC 码），待识别的码是此“闭集”中的一个。该问题可通过奇偶校验验证[8~9]来解决。若没有这样的先验知识，则称为开集识别，待识别的 LDPC 码是“无数”可能性中的一种。LDPC 码的码长通常较长，且仅由一个稀疏校验矩阵来定义，这使传统的编码识别方法纷纷失效[1]，该问题成为信道编码识别中的一个难点。

即使在无误码条件下，如何求解其稀疏校验矩阵也是一个值得研究的问题。文献[10]通过对非稀疏校验矩阵行向量的线性组合进行有限穷举来重建稀疏校验矩阵。但该方法仅适用于具有对角结构的校验矩阵，实际中的 LDPC 码并不一定符合这种条件；该方法对接收数据量有很苛刻的要求。在有误码的条件下，文献[11]在文献[10]的基础上，通过从接收数据中挑选无误码的码字来完成识别，这要求接收数据中包含大量的正确码字，因此，该方法仅具有非常弱的抗误码能力。文献[12]利用了一种寻找小重量码字的算法（称为 Canteaut-Chabaud 算法），但其应用于实际 LDPC 码时，抗误码能力依然很弱。文献[13]则采用了一种折中的穷举方法，由于运算复杂度的限制，仅适用于稀疏校验矩阵行重量很小（不大于 8）的情况。文献[14]从信息论的角度推导了实现 LDPC 码识别所需的数据量下限，但其给出的识别方法仍是穷举，其运算量对于实际 LDPC 码而言通常是不现实的。

对于接收码向量按行排列成的矩阵 $\boldsymbol{C}$，文献[12]通过寻找 $\boldsymbol{C}$ 的列空间中的小重量码字来求解校验向量。本文提出一种新的 LDPC 码开集识别方法，它同样利用了 Canteaut-Chabaud 算法[15]，然而是以一种完全不同于文献[12]的方式。

2　问题描述

本文讨论二进制 LDPC 码。码长为 n bit、码字中信息比特个数为 k 的分组码通常记为 (n,k) 码，其码率为 $\frac{k}{n}$，码空间（由所有码字构成的线性空间）维数为 k。一个 (n,k) LDPC 码由其稀疏校验矩阵 $\boldsymbol{H}$ 来定义。该矩阵是一个 $r\times n$ 矩阵，其元素在 GF(2) 上取值，行数 r 等于或略大于 $n-k$。矩阵 $\boldsymbol{H}$ 的每一行都是一个稀疏向量，其非零元素的个数远小于 n。记它的第 i 行为 $\boldsymbol{h}_i$，即有

$$t_i = wt(\boldsymbol{h}_i) \ll n,\ \ \forall i = 1,\cdots,r \tag{1}$$

其中，$wt(\cdot)$ 表示求 Hamming 重量（即元素 1 的个数，简称重量）。可见，$\boldsymbol{H}$ 整体上也是稀疏的。对于规则 LDPC 码，有 $t_1 = \cdots = t_r$；非规则 LDPC 码不一定满足此约束。本文记 $t = \max_i(t_i)$。

记此 LDPC 码的生成矩阵为 $\boldsymbol{G}$，码空间为 $\mathbb{C}$，则 $\boldsymbol{G}$ 是一个 $k\times n$ 矩阵，由 $\mathbb{C}$ 中的 k 个线性无关向量构成。用 $R(\boldsymbol{X})$ 表示由矩阵 $\boldsymbol{X}$ 的行向量张成的线性空间，则有

$$R(\boldsymbol{H}) = \mathbb{C}^{\perp} = R(\boldsymbol{G})^{\perp} \tag{2}$$

其中，$\mathbb{C}^{\perp}$ 表示 $\mathbb{C}$ 在 GF(2)^n 中的正交补空间，或称对偶空间（dual space），其维数为 $n-k$。也就是说，与一般的分组码一样，LDPC 码的生成矩阵和校验矩阵正交，即

$$\boldsymbol{G}\boldsymbol{H}^{\mathrm{T}} = \boldsymbol{O}_{k\times r} \tag{3}$$

其中，$\boldsymbol{O}_{k\times r}$ 为 $k\times r$ 全零矩阵；若用 $\boldsymbol{u}$ 表示码空间 $\mathbb{C}$ 中的任意一个码字，则有

$$\boldsymbol{u}\boldsymbol{h}_i^{\mathrm{T}} = 0,\ \ \forall i = 1,\cdots,r \tag{4}$$

本文中 GF(2)域中元素之间的运算都是该域中的运算，即乘法和模 2 加法。

由式(4)可见，$\boldsymbol{H}$ 的每一行 $\boldsymbol{h}_i$ 都定义了码字 $\boldsymbol{u}$ 所应满足的一个奇偶校验关系，称 $\boldsymbol{h}_i$ 为该 LDPC 码的稀疏校验向量。稀疏校验矩阵对于 LDPC 码的译码具有至关重要的作用。事实上，它不仅具有稀疏性，还具有其他一些性质，如应尽量避免短环[6]。这些性质使基于 $\boldsymbol{H}$ 的置信传播译码算法可达到逼近香农限的优良性能[6,16]。

LDPC 码的开集识别，就是要根据接收到的编码序列，识别出稀疏校验矩阵，为以后的译码、恢复信息序列提供条件。为简单起见，在开集识别问题中，假设已经识别出码长 n 和同步参数，从而可以将接收到的硬判决编码序列切分得到 M 个完整的码向量，记为 $\boldsymbol{c}_1,\boldsymbol{c}_2,\cdots,\boldsymbol{c}_M$，其中，$\boldsymbol{c}_i = (c_{i1},c_{i2},\cdots,c_{in})$，$c_{ij}\in$ GF(2)。由于考察硬判决序列，本文使用二进制对称信道（BSC）模型。

通常所采取的策略是分别识别出各个稀疏校验向量 $\boldsymbol{h}_i$，达到重建 $\boldsymbol{H}$ 的目的。若已知各 $\boldsymbol{h}_i$ 的最大重量 t，则需要从所有 $\sum_{i=2}^{t}\mathrm{C}_n^i$ 个稀疏向量（无需考虑重量为 1 的稀疏向量）中找出 r 个正确的校验向量。实际中 n 一般较大，如当 $n=1\,000$、$t=10$ 时，则有

$\sum_{i=2}^{t}\mathrm{C}_n^i \approx 2.7\times10^{23}$。可见，用穷举的方法来寻找稀疏校验向量是不现实的，需要通过更快速的手段来寻找。下面，针对接收码向量中无误码和有误码的情况分别进行阐述。

3 无误码条件下 LDPC 码的开集识别

3.1 识别算法

将接收到的 M 个码向量按行排列成 $M\times n$ 矩阵 $\boldsymbol{C}=[\boldsymbol{c}_1^{\mathrm{T}},\cdots,\boldsymbol{c}_M^{\mathrm{T}}]^{\mathrm{T}}$。用集合 $S=\{1,\cdots,n\}$ 表示 $\boldsymbol{C}$ 的所有列标号的集合，对于 S 的任一子集 I，可将 $\boldsymbol{C}$ 表示成关于 I 的分解形式 $\boldsymbol{C}=(\boldsymbol{V},\boldsymbol{W})_I$，其中，$\boldsymbol{V}=(\boldsymbol{c}^i)_{i\in I}$，$\boldsymbol{W}=(\boldsymbol{c}^i)_{i\in S\setminus I}$，$\boldsymbol{c}^i$ 是矩阵 $\boldsymbol{C}$ 的第 i 列。对 $\boldsymbol{C}$ 做行变换，则可得到系统形式（systematic）的矩阵

$$\boldsymbol{G}'_{\mathrm{sys}}=\left(\boldsymbol{I}_{k'},\boldsymbol{P}\right)_{I'} \tag{5}$$

其中，$k'(k'\leqslant k)$为 $\boldsymbol{C}$ 的秩，即接收到线性无关码字的个数；$\boldsymbol{I}_{k'}$ 表示 k' 阶单位阵，$\boldsymbol{P}$ 为 $k'\times(n-k')$ 矩阵；I' 为子矩阵 $\boldsymbol{I}_{k'}$ 各列在 $\boldsymbol{G}'_{\mathrm{sys}}$ 中列标号的集合。进而可以构造矩阵

$$\boldsymbol{H}'_{\mathrm{sys}}=\left(\boldsymbol{P}^{\mathrm{T}},\boldsymbol{I}_{n-k'}\right)_{I'}=\left(\boldsymbol{I}_{n-k'},\boldsymbol{P}^{\mathrm{T}}\right)_{S\setminus I'} \tag{6}$$

于是有 $\boldsymbol{G}'_{\mathrm{sys}}\boldsymbol{H}'^{\mathrm{T}}_{\mathrm{sys}}=\boldsymbol{O}_{k'\times(n-k')}$。

接收硬判决码向量中不含误码，则每一个接收码向量都是合法的码字。当 $k'=k$ 时（此时 $M\geqslant k$），显然有 $R(\boldsymbol{G}'_{\mathrm{sys}})=\mathbb{C}$ 和 $R(\boldsymbol{H}'_{\mathrm{sys}})=\mathbb{C}^{\perp}$，即 $\boldsymbol{G}'_{\mathrm{sys}}$ 和 $\boldsymbol{H}'_{\mathrm{sys}}$ 为待求编码的一对生成矩阵和校验矩阵。对于一些简单分组码的识别，至此即已完成识别任务。但是，$\boldsymbol{H}'_{\mathrm{sys}}$ 通常不是稀疏矩阵，无法用于 LDPC 码的译码。因此，对于 LDPC 码的识别，还有待进一步重建稀疏校验矩阵 $\boldsymbol{H}$，而这并没有直接或显而易见的解决方法。此外，实际中接收码字的数量有可能很有限，不能保证 $M\geqslant k$ 或 $k'=k$。当 $k'<k$ 时，有

$$\begin{cases}R(\boldsymbol{G}'_{\mathrm{sys}})\subset\mathbb{C}\\ \mathbb{C}^{\perp}\subset R(\boldsymbol{H}'_{\mathrm{sys}})=R(\boldsymbol{G}'_{\mathrm{sys}})^{\perp}\end{cases} \tag{7}$$

即完整的码空间 $\mathbb{C}$ 或对偶空间 $\mathbb{C}^{\perp}$ 都未知。本文需要研究能否利用数量有限的接收码字，尽可能快地找出所有稀疏校验向量 $\boldsymbol{h}_i$。可见，即使在无误码条件下，LDPC 码的开集识别问题也是一个值得深入研究的问题。

在公开发表的文献中，文献[10]首次阐述了该问题，并提出了行间线性组合有限穷举算法，将矩阵 $\boldsymbol{H}'_{\mathrm{sys}}$ 稀疏化得到 $\boldsymbol{H}$。由于现有的有误码条件下 LDPC 码识别算法[11~14]不能充分利用“无误码”的特殊优势，无法高效地实现无误码条件下的识别，因此，文献[10]算法是目前针对此问题的仅有算法。然而，该算法存在 2 个方面的局限。首先，它主要适用于稀疏校验矩阵 $\boldsymbol{H}$ 具有“双对角”或“多对角”结构的情况，不具有通用性；其次，它要求接收到 k 个线性无关的码字。事实上，文献[14]已经从信息论的角度证明，只需要 $O(t\mathrm{lb}n)$ 数量的码字就能够实现 LDPC 码的识别，而不需要 k 个码字。对于这一结论，下面给出一种更为简单、直观的解释。

长度为 n、重量不超过 t 且不是该 LDPC 码稀疏校验向量的向量个数为 $a(t)=\sum_{i=2}^{t}\mathrm{C}_n^i-r$，每一个这样的向量与任一非零接收码字正交（即满足式(4)）的概率为 $\frac{1}{2}$。要求这些向量中与所有接收码字都正交的向量个数的期望远小于 1，也就是 $\frac{a(t)}{2^{k'}}\ll1$；或者，这些向量中不存在与所有接收码字都正交的向量的概率应趋近于 1，即 $\left(1-\frac{1}{2^{k'}}\right)^{a(t)}\to1$。这 2 种表述是等价的，它们都等价于

$$2^{k'}\gg a(t)=\sum_{i=2}^{t}\mathrm{C}_n^i-r \tag{8}$$

使用数学归纳法可以证明如下不等式

$$\sum_{i=2}^{t}\mathrm{C}_n^i<\frac{n^t}{t!},\ \forall n,2\leqslant t<\frac{n}{2} \tag{9}$$

实际上，为了保证 LDPC 码的纠错能力，t 不宜过小，通常有 $t\geqslant5$。于是，可认为 $\frac{n^t}{t!}\ll n^t$，结合式(9)可知，取 $2^{k'}\geqslant n^t$ 即可满足式(8)的要求。也就是说，当接收码字的数量 M 满足

$$M\geqslant k'\geqslant t\,\mathrm{lb}\,n \tag{10}$$

利用式(4)进行验证，就可以足够高的概率排除所有不是校验向量的稀疏向量，而仅保留该 LDPC 码的稀疏校验向量。如当 $n=1\,000$、$t=8$ 且码率为 $\frac{1}{2}$ 时，理论上仅需要 80 个接收码字（而非 $k=500$ 个）即可完成识别。文献[14]给出的是当 $n\to\infty$ 时，数据量

的渐近下界。式(10)所给出的数据量下界，对于实际中的 LDPC 码，则能以足够高的概率保证空间 $R(\boldsymbol{H}'_{\text{sys}})$ 中所有重量不大于 t 的向量都是该码的稀疏校验向量。

为了克服文献[10]方法的局限性，本文提出使用 Canteaut-Chabaud 算法[15]来识别 LDPC 码的稀疏校验向量。对于任意矩阵 $\boldsymbol{X}$（其元素取自 GF(2)），该算法能以迭代的方式快速找出 $R(\boldsymbol{X})$ 中的小重量向量，它原本用于寻找码空间中的小重量码字。这里，本文利用它寻找 $R(\boldsymbol{H}'_{\text{sys}})$ 中的小重量向量，即待识别 LDPC 码的稀疏校验向量 $\boldsymbol{h}_i$，$i=1,\cdots,r$。原算法需要事先给定待求向量的最大重量 t，且找出一个小重量向量即结束；而这里最大重量 t 未知，且需要找出 r 个小重量向量。因此，需要对原算法进行如下修改：一是在寻找过程中逐步确定 t 的值，二是采用新的结束准则。无误码条件下，基于 Canteaut-Chabaud 算法的 LDPC 码开集识别算法流程如下。

初始化　按照式(6)构造矩阵 $\boldsymbol{H}'_{\text{sys}}=(\boldsymbol{I}_{n-k'},\boldsymbol{P}^{\text{T}})_I$（其中，初始列标号集合 $I=S\setminus I'$），重量阈值 $t'=\left\lfloor \dfrac{k'}{\text{lb}\,n} \right\rfloor$，空间 $R(\boldsymbol{H}'_{\text{sys}})$ 中小重量向量的集合 $\boldsymbol{\Phi}=\varnothing$，计数器置零。

1) 计数器加 1；将 I 随机分成 2 个子集 I_1 和 I_2，分别包含 $\left\lfloor \dfrac{n-k'}{2} \right\rfloor$ 和 $\left\lceil \dfrac{n-k'}{2} \right\rceil$ 个元素；将 I_1 和 I_2 视为矩阵 $\boldsymbol{H}'_{\text{sys}}$ 行标号的集合，矩阵 $\boldsymbol{H}'_{\text{sys}}$ 则分为 $\boldsymbol{H}_1$ 和 $\boldsymbol{H}_2$ 2 部分，分别由行标号属于 I_1 和 I_2 的行构成。

2) 随机选取元素个数为 σ 的列标号集合 L，满足 $L\subset J\overset{\text{def}}{=}S\setminus I$。

3) 计算 $\boldsymbol{H}_1$ 中任意 p 行的和向量 $\boldsymbol{s}_1$，将其在 L 上的取值 $\boldsymbol{s}_{1|L}$ 记录在表 Δ_1 中；计算 $\boldsymbol{H}_2$ 中任意 p 行的和向量 $\boldsymbol{s}_2$，将其在 L 上的取值 $\boldsymbol{s}_{2|L}$ 记录在表 Δ_2 中。

4) 根据表 Δ_1 和 Δ_2，考察所有满足 $\boldsymbol{s}_{1|L}=\boldsymbol{s}_{2|L}$ 的 $(\boldsymbol{s}_1,\boldsymbol{s}_2)$ 组合，若有 $wt(\boldsymbol{s}_{1|J\setminus L}+\boldsymbol{s}_{2|J\setminus L})\leqslant t'-2p$，且有 $\boldsymbol{h}=\boldsymbol{s}_1+\boldsymbol{s}_2\notin\boldsymbol{\Phi}$，则有：

① 令 $\boldsymbol{\Phi}=\boldsymbol{\Phi}\cup\{\boldsymbol{h}\}$，计数器置零；

② 若 $t'\geqslant 2wt(\boldsymbol{h})-2$，则令 $t'=2wt(\boldsymbol{h})-3$，然后令 $\boldsymbol{\Phi}=\boldsymbol{\Phi}\setminus\{\boldsymbol{h}'\,|\,wt(\boldsymbol{h}')>t'\}$。

5) 若计数器数值达到正整数 T，算法结束；否则，随机选取 $\eta\in I$ 和 $\mu\in J$，记集合 $\hat{I}=(I\setminus\{\eta\})\cup\{\mu\}$，并通过行变换将矩阵 $\boldsymbol{H}'_{\text{sys}}$ 化为 $\hat{\boldsymbol{H}}'_{\text{sys}}=(\boldsymbol{I}_{n-k'},\hat{\boldsymbol{P}}^{\text{T}})_{\hat{I}}$，令 $I=\hat{I}$ 且 $\boldsymbol{H}'_{\text{sys}}=\hat{\boldsymbol{H}}'_{\text{sys}}$，回到步骤 1)。

按照步骤 4)，每当找到一个向量 $\boldsymbol{h}$，若有 $t'\geqslant 2wt(\boldsymbol{h})-2$，就将阈值 t' 更新为 $2wt(\boldsymbol{h})-3$。这是由于 2 个重量为 $wt(\boldsymbol{h})$ 的稀疏校验向量之和仍是该 LDPC 码的校验向量，其重量的最小值为 $2wt(\boldsymbol{h})-2$（假设不存在长度为 4 的环）。步骤 4)的做法既可以避免找出这样的校验向量，同时也适应于非规则 LDPC 码的要求，可以找出重量在一定范围内的所有稀疏校验向量。当算法结束时，找出的所有小重量向量存于集合 $\boldsymbol{\Phi}$ 中，它们即被视为待识别 LDPC 码的稀疏校验向量。

上述算法的参数包括 σ、p 和 T。参数 σ 和 p 的选取方法已由文献[15]给出，并将在下一节进行讨论。参数 T 的意义是：若在连续 T 次迭代中没有找到新的小重量向量，则算法结束。

3.2　迭代次数分析

为了设置参数 T 的值，或分析算法的运算量，都需要对算法的迭代次数进行分析。

用随机变量 N_i 表示算法找出某个稀疏校验向量 $\boldsymbol{h}_i$ 所需迭代次数，$\forall i=1,\cdots,r$。由于 $\boldsymbol{h}_i$ 的稀疏性和其元素 1 所在位置的随机性，各变量 N_i 可视为相互独立。记 $\boldsymbol{h}_i$ 中元素 1 所在位置的集合为 S_i，算法每次迭代的结果取决于当前列标号集合 I 与 S_i 之间的重合关系。若这一重合关系在初始化时是随机的，则在每次迭代中也可视为完全随机的。因此，可认为迭代次数 N_i 具有“无记忆性”，即对任意正整数 j_1 和 j_2（$j_1>j_2$），有 $\Pr(N_i>j_1\,|\,N_i>j_2)=\Pr(N_i>j_1-j_2)$。“无记忆性”是指数分布的固有性质。通常，指数分布用来描述连续型随机变量，但相关的结论可以推广应用于离散型随机变量的情形。本文考虑使用参数为 λ_i 的指数分布来近似描述 N_i 的行为，则其期望为 $\text{E}(N_i)=\dfrac{1}{\lambda_i}$，方差为 $V(N_i)=\dfrac{1}{\lambda_i^2}$。文献[15]给出了 $\text{E}(N_i)$、$V(N_i)$ 的计算方法，其结果满足 $\text{E}(N_i)\approx\sqrt{V(N_i)}$，于是参数 λ_i 的取值为 $\lambda_i=\dfrac{1}{\sqrt{V(N_i)}}$。对于 2 个不同的稀疏校验向量 $\boldsymbol{h}_i$ 和 $\boldsymbol{h}_j$，如果它们的重量满足 $t_i=t_j$，则有 $V(N_i)=V(N_j)$，于是有 $\lambda_i=\lambda_j$。

首先讨论算法参数 T 的取值。当还剩 s 个稀疏

校验向量未找到时，找到下一个稀疏校验向量所需迭代次数可表示为 $N_{\min}=\min\limits_{j=1,\cdots,s}(N_{i_j})$，其中，$i_j\in\{1,\cdots,r\}$。多个独立指数分布随机变量的最小值仍服从指数分布。有 $V(N_{\min})=\mathrm{E}(N_{\min})^2$，且

$$\mathrm{E}(N_{\min})=\left(\sum_{j=1}^{s}\lambda_{i_j}\right)^{-1} \tag{11}$$

为了尽量保证算法在找到下一个稀疏校验向量之前不结束，可使用“3 倍标准差”准则来设置 T。由于 s 未知，因此，将 T 设置为

$$\begin{aligned}T&=\left\lceil\max_{s,\{i_1,\cdots,i_s\}}\left(\mathrm{E}(N_{\min})+3\sqrt{V(N_{\min})}\right)\right\rceil\\&=\left\lceil\max_{s,\{i_1,\cdots,i_s\}}\left(4\mathrm{E}(N_{\min})\right)\right\rceil\\&=\left\lceil 4\max_{i=1,\cdots,r}\left(\lambda_i^{-1}\right)\right\rceil\\&=\left\lceil 4\max_{i=1,\cdots,r}\left(\sqrt{V(N_i)}\right)\right\rceil\end{aligned} \tag{12}$$

实际上，在步骤 4）中更新 t' 值的同时，也根据已找出的小重量向量来更新参数 T。

下面，分析算法的总迭代次数。设算法获取正确阈值 t' 所需迭代次数可忽略不计，且算法结束前已找出所有 r 个稀疏校验向量，则总迭代次数可以表示为 $N=N_{\max}+T$，其中，$N_{\max}=\max\limits_{i=1,\cdots,r}(N_i)$。多个指数分布随机变量的最大值不再服从指数分布，但可推导其期望和二阶矩分别为

$$\mathrm{E}\left(N_{\max}\right)=\sum_{i=1}^{r}\left[\lambda_i^{-1}+\lambda_i\sum_{s=1}^{r-1}\left((-1)^s\sum_{J_{s\backslash i}}\left(\lambda_i+\sum_{j\in J_{s\backslash i}}\lambda_j\right)^{-2}\right)\right] \tag{13}$$

$$\mathrm{E}\left(N_{\max}^2\right)=2\sum_{i=1}^{r}\left[\lambda_i^{-2}+\lambda_i\sum_{s=1}^{r-1}\left((-1)^s\sum_{J_{s\backslash i}}\left(\lambda_i+\sum_{j\in J_{s\backslash i}}\lambda_j\right)^{-3}\right)\right] \tag{14}$$

其中，集合 $J_{s\backslash i}$ 表示集合 $\{1,\cdots,r\}\backslash\{i\}$ 的任意包含 s 个元素的子集。进而，可计算方差 $V(N_{\max})=\mathrm{E}(N_{\max}^2)-\mathrm{E}(N_{\max})^2$。对于规则 LDPC 码，所有稀疏校验向量重量相等，故有 $\lambda_1=\cdots=\lambda_r$；对于非规则 LDPC 码，其稀疏校验向量的重量通常也仅有少量不同取值，则 $\lambda_1,\cdots,\lambda_r$ 也仅有少量不同取值。因此，式(13)和式(14)通常可以化简，如对于规则 LDPC 码，有[17]

$$\mathrm{E}\left(N_{\max}\right)=\frac{r}{\lambda}\left[1+\sum_{s=1}^{r-1}(-1)^s\mathrm{C}_{r-1}^{s}(s+1)^{-2}\right]=\frac{1}{\lambda}\sum_{s=1}^{r}s^{-1} \tag{15}$$

$$\begin{aligned}V\left(N_{\max}\right)&=\frac{2r}{\lambda^2}\left[1+\sum_{s=1}^{r-1}(-1)^s\mathrm{C}_{r-1}^{s}(s+1)^{-3}\right]-\mathrm{E}\left(N_{\max}\right)^2\\&=\frac{1}{\lambda^2}\sum_{s=1}^{r}s^{-2}\end{aligned} \tag{16}$$

其中，$\lambda=\lambda_i$，$\forall i$。算法总迭代次数的期望和方差分别为

$$\mathrm{E}(N)=\mathrm{E}\left(N_{\max}\right)+T \tag{17}$$

$$V(N)=V(N_{\max}) \tag{18}$$

对于 LDPC 码，最大重量 t 一般较小，根据文献[15]，这种情况下通常设参数 $p=1$；为了尽量减少算法步骤 4)的运算量，设参数 $\sigma=\lceil\mathrm{lb}n\rceil$。于是，算法每次迭代的运算量为 $W\approx k'(n-k')+\sigma\left(\frac{n-k'}{2}\right)^2$ 次二进制运算。算法的平均总运算量为 $\mathrm{E}(N)W$ 次二进制运算。

3.3 在 QC-LDPC 码情况下的改进

QC-LDPC 码[16]不仅方便构造，而且能够实现高效编、译码，因而在各类协议中被广泛采用。对于 QC-LDPC 码，码长 $n=ml$，稀疏校验矩阵行数 $r=mz$，其中，m、l、z 都是整数，并满足如下性质：如果向量 $\boldsymbol{h}=[\boldsymbol{h}_{(1)},\cdots,\boldsymbol{h}_{(l)}]$ 是该码的校验向量（$\forall i=1,\cdots,l$，$\boldsymbol{h}_{(i)}$ 长度为 m，表示 $\boldsymbol{h}$ 的第 i 段），则向量 $\boldsymbol{h}^1=[\boldsymbol{h}_{(1)}^1,\cdots,\boldsymbol{h}_{(l)}^1]$ 也是其校验向量，其中，$\boldsymbol{h}_{(i)}^1$ 表示由 $\boldsymbol{h}_{(i)}$ 循环右移 1 位的结果。可见，所有 r 个稀疏校验向量可以分为 z 组，每组包含 m 个向量；同组的所有向量可由其中任一向量通过不断地分段循环移位得到。根据 QC-LDPC 码的这一特点，识别算法可以做如下改进。

考察码长 n 的所有因数分解组合 (m,l)，当算法找出第一个校验向量 $\boldsymbol{h}$（不一定是待求的稀疏校验向量）时，按照每一组 (m,l) 数值对 $\boldsymbol{h}$ 进行分段循环移位；若分段循环移位的结果与所有接收码字正交（即满足式(4)），说明此时的 (m,l) 数值正确。以此方法得到正确的 m 和 l 后，算法每找出一个向量，即把相应的一组（m 个）向量加入集合 $\boldsymbol{\Phi}$ 中；若连续 T^{QC} 次迭代没有找到新的向量，算法结束。算法的其他操作不变。

改进后的算法找到第 i ($i=1,\cdots,z$)组稀疏校验向量所需迭代次数为 $N_{[i]}=\min\limits_{j=1,\cdots,m}(N_{(i-1)m+j})$。由于同组稀疏校验向量的重量相同，故指数分布随机变量 $N_{(i-1)m+j}$ ($j=1,\cdots,m$)具有相同参数，记为 $\lambda_{[i]}$。于是，

$N_{[i]}$ 服从参数为 $m\lambda_{[i]}$ 的指数分布。可知参数 T^{QC} 应设置为

$$T^{\mathrm{QC}}=\left\lceil 4\max_{i=1,\cdots,z}\left(\left(m\lambda_{[i]}\right)^{-1}\right)\right\rceil=\left\lceil\frac{4}{m}\max_{j=1,\cdots,r}\left(\sqrt{V(N_j)}\right)\right\rceil=\left\lceil\frac{T}{m}\right\rceil \tag{19}$$

改进后算法的总迭代次数为 $N^{\mathrm{QC}}=N_{\max}^{\mathrm{QC}}+T^{\mathrm{QC}}$，其中，$N_{\max}^{\mathrm{QC}}=\max_{i=1,\cdots,z}(N_{[i]})$。$N_{\max}^{\mathrm{QC}}$ 的期望和二阶矩的形式与式(13)和式(14)相似，且容易证明

$$\mathrm{E}\left(N_{\max}^{\mathrm{QC}}\right)<\frac{1}{m}\mathrm{E}\left(N_{\max}\right) \tag{20}$$

$$\mathrm{E}\left[\left(N_{\max}^{\mathrm{QC}}\right)^2\right]<\frac{1}{m^2}\mathrm{E}\left(N_{\max}^2\right) \tag{21}$$

于是，结合式(17)和式(19)有

$$\mathrm{E}(N^{\mathrm{QC}})=\mathrm{E}\left(N_{\max}^{\mathrm{QC}}\right)+T^{\mathrm{QC}}<\frac{\mathrm{E}(N)}{m} \tag{22}$$

即算法在改进之后，平均总迭代次数将大大降低至原来的 $\frac{1}{m}$ 以下。对于规则 QC-LDPC 码，$N_{\max}^{\mathrm{QC}}$ 的期望和方差分别为

$$\mathrm{E}\left(N_{\max}^{\mathrm{QC}}\right)=\frac{1}{m\lambda}\sum_{i=1}^{z}i^{-1} \tag{23}$$

$$V\left(N_{\max}^{\mathrm{QC}}\right)=\frac{1}{m^2\lambda^2}\sum_{i=1}^{z}i^{-2} \tag{24}$$

4　有误码条件下 LDPC 码的开集识别

在有误码的条件下，接收码向量 $\boldsymbol{c}_1,\boldsymbol{c}_2,\cdots,\boldsymbol{c}_M$ 不再是 LDPC 码的正确码字，式(7)不再成立，即空间 $R(\boldsymbol{H}'_{\mathrm{sys}})$ 中不包含完整的对偶空间 $\mathcal{C}^{\perp}$。记 BSC 信道的比特错误概率为 p，则待识别 LDPC 码的任一稀疏校验向量 $\boldsymbol{h}_i$ 与任一接收码向量正交的概率 p_{t_i} [14]为

$$p_{t_i}=\frac{1+(1-2p)^{t_i}}{2} \tag{25}$$

用随机变量 X_i 表示 M 个接收码向量中，与 $\boldsymbol{h}_i$ 正交的码向量个数，则 X_i 服从二项分布，即

$$\Pr\left(X_i=j\right)=\mathrm{C}_M^j p_{t_i}^j\left(1-p_{t_i}\right)^{M-j},\quad j=0,1,\cdots,M \tag{26}$$

对于 LDPC 码，$X_1,\cdots,X_r$ 可视为相互独立。当且仅当 $X_i=M$，即 $\boldsymbol{h}_i$ 与所有接收码向量正交时，有 $\boldsymbol{h}_i\in R(\boldsymbol{H}'_{\mathrm{sys}})$，其概率为 $\Pr(X_i=M)=p_{t_i}^M$。当 p 固定时，接收码向量越多（即 M 越大），此概率越小。当 M 较大时，常有 $\boldsymbol{h}_i\notin R(\boldsymbol{H}'_{\mathrm{sys}})$，$\forall i$，在这种情况下，无法从 $R(\boldsymbol{H}'_{\mathrm{sys}})$ 中找到稀疏校验向量。

为了解决这一问题，从 M 个接收向量中随机挑选 M_c 个，按行排列成矩阵 $\boldsymbol{C}_c$。对 $\boldsymbol{C}_c$ 做行变换并按式(6)的方式构造得到矩阵 $\boldsymbol{H}'_{c,\mathrm{sys}}$。由于 M_c 可取足够小的值（根据式(10)，最小能取到 $t\,\mathrm{lb}\,n$），则 $\boldsymbol{h}_i\in R(\boldsymbol{H}'_{c,\mathrm{sys}})$ 成立的概率 $p_{t_i}^{M_c}$ 较大。

设置正整数 M_c 以及 $N_{c,1}$、$N_{c,2}$，当接收到 M 个有误码的码向量时，LDPC 码的开集识别方法如下。

初始化　稀疏校验向量集合 $\boldsymbol{\Phi}=\varnothing$，计数器置零。

1) 计数器加 1，从 M 个码向量中随机挑选 M_c 个，构造矩阵 $\boldsymbol{H}'_{c,\mathrm{sys}}$。

2) 利用 3.1 节算法寻找空间 $R(\boldsymbol{H}'_{c,\mathrm{sys}})$ 中重量不大于 t 的向量（迭代次数固定为 $N_{c,1}$ 次），将找到的向量加入集合 $\boldsymbol{\Phi}$。

3) 若计数器达到 $N_{c,2}$，结束；否则回到步骤 1)。

参数 M_c 的取值是此方法的关键。为了推导 M_c，简单起见，这里，考虑规则 LDPC 码。则 $\forall i$，$t_i=t$，并记 $p_{t_i}=p_t$。于是，空间 $R(\boldsymbol{H}'_{c,\mathrm{sys}})$ 中稀疏校验向量个数的期望为 $rp_t^{M_c}$。对任意 $\boldsymbol{h}_i\in R(\boldsymbol{H}'_{c,\mathrm{sys}})$，利用 3.1 节算法找到 $\boldsymbol{h}_i$ 所需的迭代次数记为 $N_i(M_c)$，服从参数为 $\lambda(t,M_c)$ 的指数分布。找出任一稀疏校验向量所需迭代次数 $N_{\min}(M_c)$ 为这些指数分布随机变量 $N_i(M_c)$ 的最小值，其期望为

$$\mathrm{E}\left(N_{\min}(M_c)\right)=\frac{1}{rp_t^{M_c}\lambda(t,M_c)} \tag{27}$$

为了能在步骤 2)中尽快找到所求向量，参数 M_c 的取值应使式(27)取得最小值，即

$$M_c=\arg\max_{X:\ t\,\mathrm{lb}\,n\leqslant X\leqslant xM}\left[p_t^X\lambda(t,X)\right] \tag{28}$$

其中，$x<1$，设置 x 的目的是为了步骤 1)挑选码向量时具有较大的自由度。式(28)与 r 无关，但与 t 及 BSC 信道的错误概率 p 有关。因此，本节方法需要假定 t、p 已知。

对于非规则 LDPC 码，有 $t_i\leqslant t$，$\forall i$。为了保证非规则码的优良性能，在实际设计时通常使各稀疏校验向量的重量相近，即有 $t_i\approx t$，$\forall i$。在这种情况下，仍可利用式(28)来近似计算 M_c 的值。

步骤 2)采用 3.1 节的算法，但做以下改动：由于 t 已知，可去掉其中更新阈值 t' 的步骤；为了节省运算量，将它的结束准则改为固定执行 $N_{c,1}$ 次迭代后结束。本节方法在每次迭代中挑选 M_c 个码向量来展开识别，其迭代次数 $N_{c,2}$ 可根据实际中对运算量的限制来设定。

在 QC-LDPC 码的情况下，仍可对本节方法进行改进。其方法与 3.3 节所述大致相同，但在利用找到的向量 $\boldsymbol{h}$ 来识别参数 m 、l 时，由于存在误码，故需借助门限才能判断 $\boldsymbol{h}$ 分段循环移位后是否仍是校验向量。此门限的设置可参考文献[14,18]等，不再赘述。

5 仿真实验

针对本文提出的无误码和有误码条件下 LDPC 码的开集识别算法，进行了大量的仿真实验。所使用的编码为 IEEE 802.11n 标准[7]中定义的各种 LDPC 码（码长可选为 648 bit、1 296 bit、1 944 bit 等，码率有 $\frac{1}{2}$、$\frac{2}{3}$、$\frac{3}{4}$、$\frac{5}{6}$ 等），以及一种随机构造的（1 008，504）规则 QC-LDPC 码。本节给出实验结果，并分别与已有算法的相应实验结果进行对比分析。

5.1 无误码条件下的识别

在无误码条件下，5 组不同的实验条件如表 1 所示。实验使用的 LDPC 码为(648, 324)码和(1 008, 504)码。前者的稀疏校验矩阵 $\boldsymbol{H}$ 具有双对角结构，本文考虑对角结构在 $\boldsymbol{H}$ 中的位置已知或未知，以及接收独立码字个数 $k'=k$ 或 $k'<k$ 的情况；后者 $\boldsymbol{H}$ 不具有对角结构。表 2 给出了在这 5 组实验条件下，本文算法与文献[10]算法识别结果的对比。根据 3.2 节，本文算法参数为 $p=1$，$\sigma=10$，T（稳定后的值）在表 2 中列出。2 种算法分别进行蒙特卡洛实验，平均每次实验中识别出的稀疏校验向量个数占总数 r 的百分比称为平均识别率，记为 $\bar{P}$；算法的平均迭代次数记为 $\bar{N}$。由表 2 可知，只有当 $\boldsymbol{H}$ 具有对角结构、对角结构位置已知，且 $k'=k$ 时，文献[10]算法才能 100%识别成功；本文算法则没有这些限制，均能完成识别。当 2 种算法都 100%识别成功时，根据各自的迭代次数和每次迭代的运算量可知，文献[10]算法的运算量约为 $12r^2n\approx 8.2\times10^8$ 次二进制运算，而本文算法约需 $4185W\approx 1.5\times10^9$ 次二进制运算。应当指出的是，若扣除结束前的最后 T 次迭代，本文算法找出全部稀疏校验向量实际所用的平均迭代次数为 $4185-T$ 次，仅相当于 5.7×10^8 次二进制运算。

表 1 无误码条件下进行识别的 5 组实验条件

实验条件	(n,k)	对角结构	对角位置	k'
第 1 组	(648, 324)	双对角	已知	324
第 2 组	(648, 324)	双对角	未知	324
第 3 组	(648, 324)	双对角	已知	200
第 4 组	(1 008, 504)	—	—	504
第 5 组	(1 008, 504)	—	—	300

表 2 无误码条件下本文算法与文献[10]算法识别结果对比

实验条件	文献[10]算法		本文算法		
	$\bar{P}$	$\bar{N}$	$\bar{P}$	$\bar{N}$	T
第 1 组	100%	12	100%	4 185	2 630
第 2 组	33.5%	27	100%	6 748	2 630
第 3 组	0.5%	10	100%	42 026	16 580
第 4 组	5.3%	12	100%	4 986	1 854
第 5 组	0.2%	9	100%	18 182	6 799

为了测试接收数据量 M 对本文算法的影响，在不同数据量情况下对码率为 $\frac{2}{3}$ 的(648, 432)码进行了实验。该码为 QC-LDPC 码（$m=27$，$l=24$），各稀疏校验向量重量均为 11。实验中，令归一化数据量 $\frac{M}{n}$ 在 $0.3\sim\frac{2}{3}$ 间变化（满足式(10)的要求），分别使用 3.1 节算法和 3.3 节的改进算法进行识别。本文 2 种算法在各种数据量条件下都实现了 100%的平均识别率。图 1 给出了 2 种算法平均迭代次数 $\bar{N}$ 随 $\frac{M}{n}$ 的变化情况，并分别与式(17)和式(22)的理论值进行了对比。首先，迭代次数的理论值与实际值十分吻合，仅当 $\frac{M}{n}$ 接近码率时改进算法的理论值略高于实际值；其次，平均迭代次数随接收数据量的减少而上升，这说明接收数据量越少，完成识别所需运算量越多；最后，针对 QC-LDPC 码的改进算法与原算法相比，所需迭代次数大为减少（如当 $\frac{M}{n}=\frac{2}{3}$ 时，平均迭代次数从 3 659 降为 58），大大提高了识别效率。针对不同码长、码率的 LDPC 码进行实验，均得到了类似的结论。

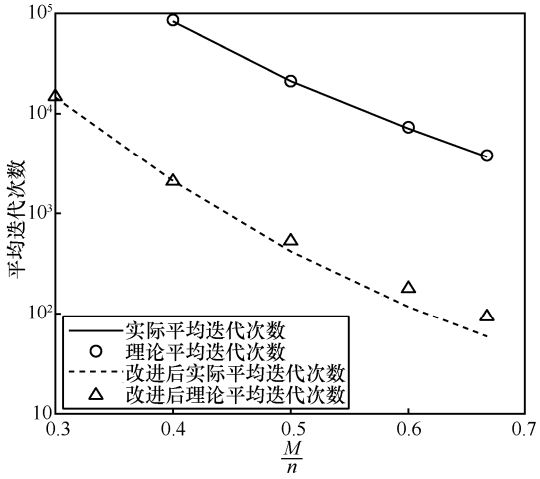


图 1　本文算法识别(648, 432)LDPC 码时，平均迭代次数 $\bar{N}$ 随归一化数据量 $\frac{M}{n}$ 的变化

记本文算法找出 i 个稀疏校验向量所需迭代次数为 $N(i)$，这里，考察 $N(i)$ 占总迭代次数 N 的百分比，即 $\frac{N(i)}{N}\times 100\%$。在上述针对(648, 432)码的实验中，这一百分比的平均值随 i（$i=1,\cdots,r$，$r=216$）的变化情况如图 2 所示。可见，本文算法一开始总能以较少的迭代次数迅速找出大量的稀疏校验向量，而找出剩下的少数稀疏校验向量则须耗费较多的迭代。一方面，这是由迭代次数的指数分布特性决定的；另一方面，由于 r 未知，算法在结束前不得不“浪费” T（或 T^{QC}）次迭代来确保已得到全部待求向量。对于本文针对 QC-LDPC 码的改进算法，这一现象尤为明显。如在图 2 中，当 $\frac{M}{n}=0.6$ 时，改进算法找到全部稀疏校验向量时的迭代次数仅占其总迭代次数的不到 10%。

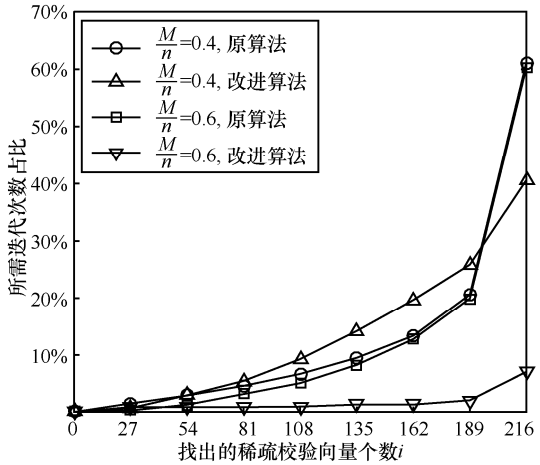


图 2　本文算法找到 i 个稀疏校验向量所需迭代次数占总迭代次数的百分比（稀疏校验向量总数为 216 个）

5.2　有误码条件下的识别

有误码条件下的 7 组不同实验条件如表 3 所示。实验针对 3 种 LDPC 码，其中，(1 008, 504)码和(648, 432)码为规则码，前者 $t=6$、$r=504$，后者 $t=11$、$r=216$；(648, 324)码为非规则码，$t=8$（t_i 为 7 或 8）、$r=324$。表 3 中列出了 BSC 信道的比特错误概率 p、接收码向量中的误码字率（含误码的码向量个数占总数的百分比）、接收码向量总数 M，以及参数 M_c。其中，第 5、7 组实验条件的 M 值满足 $M<n$，其他组中皆满足 $M=n$；除第 5 组外，各组中的 M_c 值皆根据式(28)（式中 x 取 0.5）得到，第 5 组的 M_c 值根据式(28)应为 250，但本文将其设置为 324 以便结合第 4 组实验条件，考察参数 M 对算法性能的影响。此外，实验中设置 $N_{c,1}=10\,000$。表 4 给出了各组实验条件下，本文第 4 节方法的识别结果。可见，在指定的 $N_{c,2}$ 次迭代之内，通常能识别出相应 LDPC 码的大部分稀疏校验向量。对比第 1~3 组条件下的识别结果可知，p 越小，识别效果越好；对比第 4、5 组或第 6、7 组可知，M 越大，识别效果越好。

表 3　有误码条件下进行识别的 7 组实验条件

实验条件	(n,k)	p	误码字率	M	M_c
第 1 组	(1 008, 504)	0.01	100%	1 008	60
第 2 组	(1 008, 504)	0.005	99.6%	1 008	95
第 3 组	(1 008, 504)	0.002	86.9%	1 008	200
第 4 组	(648, 432)	0.001	48.2%	648	324
第 5 组	(648, 432)	0.001	48.2%	500	324
第 6 组	(648, 324)	0.002	72.1%	648	240
第 7 组	(648, 324)	0.002	72.1%	500	240

表 4 中的识别结果显示，本文算法并未识别出全部的稀疏校验向量。然而，利用识别出的稀疏校验向量，使用和积算法[6]对接收码向量的软判决向量进行译码，可使误比特率和误码字率大为降低。这一译码结果也已在表 4 中给出。例如对第 2 组实验条件，经过 $N_{c,2}=100$ 次迭代平均可识别出 385 个稀疏校验向量（平均识别率 $\bar{P}$ 为 76.4%），用它们进行译码，可使接收码向量中的误比特率由 0.005 降至 7.5×10^{-5}，而误码字率由 99.6%降至 6.2%。利用译码后的码向量继续进行识别，有望快速找出剩下的稀疏校验向量。

在有误码条件下，文献[11]的方法仅能处理

10^{-4} 量级的信道误比特率，并要求接收码向量中包含大量正确码字。文献[12]方法处理 $n=1\,000$、$t=6$ 的 LDPC 码时，要求 p 不高于 0.002。这 2 种方法均需使用大量的接收码向量（$M>n$）。可见，本文方法的识别能力优于这 2 种方法。文献[13,14]方法由于运算复杂度的限制，通常不能应用于重量 $t>8$ 的 LDPC 码，而本文算法能够处理 t 值更高的情况，如表 3 中 $t=11$ 的(648, 432)码。

表 4　有误码条件下本文方法的识别结果和相应译码效果

实验结果	识别结果		译码效果	
	$N_{c,2}$	$\bar{P}$	p	误码字率
第 1 组	400	62.9%	6.3×10^{-4}	42.1%
第 2 组	100	76.4%	7.5×10^{-5}	6.2%
第 3 组	100	94.3%	0	0
第 4 组	100	60.7%	7.9×10^{-5}	4.8%
第 5 组	100	43.9%	3.5×10^{-4}	19.9%
第 6 组	100	69.6%	2.2×10^{-4}	12.7%
第 7 组	100	57.2%	4.5×10^{-4}	25.6%

表 3 中的 3 种编码都是 QC-LDPC 码，可采用相应的改进算法进行识别。在表 3 各组实验条件下，改进算法对稀疏校验向量的识别率达到 100%时，所需迭代次数 $N_{c,2}$ 的平均值 $\bar{N}_{c,2}$，如表 5 所示。可见，改进算法平均仅需数次迭代就能达到 100%识别率，即识别出全部的稀疏校验向量。例如对于(1 008, 504)码，其参数 $m=168$、$z=3$，故仅需从每组 168 个稀疏校验向量中找出任意 1 个即可，而本文算法由于具有如图 2 所示的特点，能快速完成这一任务。

表 5　有误码条件下本文方法针对 QC-LDPC 码改进后，识别率达到 100%时所需平均迭代次数 $\bar{N}_{c,2}$

实验条件	$\bar{N}_{c,2}$
第 1 组	5.9
第 2 组	1.5
第 3 组	1.0
第 4 组	4.6
第 5 组	5.4
第 6 组	4.8
第 7 组	5.2

6　结束语

针对目前 LDPC 码开集识别问题研究不足的情况，本文基于一种快速寻找小重量码字的算法，提出了新的 LDPC 码开集识别方法。在无误码条件下，本文算法利用接收码字实现了稀疏校验矩阵的全自动识别，克服了已有算法在校验矩阵结构、所需码字数量等方面的诸多局限。在有误码条件下，本文方法具有较强的抗误码能力和较低的运算复杂度，与已有方法相比，更能满足实际应用的需求。并且，本文指出利用 QC-LDPC 码的准循环特性对算法进行改进，可以大大提高识别效率。

在有误码条件下，本文方法需利用重量 t、误比特率 p 来设置参数 M_c。当它们未知时，可对 M_c 进行试探性取值，再根据找出的校验向量逐步进行调整，这可作为下一步的研究内容。另外，如何将 5.2 节提到的识别和译码更加完善地结合起来，以提高识别性能，也值得进一步研究。

参考文献：

[1] 解辉, 黄知涛, 王丰华. 信道编码盲识别技术研究进展[J]. 电子学报, 2013, 41(6): 1166-1176.
XIE H, HUANG Z T, WANG F H. Research progress of blind recognition of channel coding[J].Acta Electronica Sinica, 2013, 41(6): 1166-1176.

[2] 阚永红, 曾伟涛, 陈健. 基于概率逼近的本原 BCH 码编码参数的盲识别方法[J]. 电子与信息学报, 2014, 36(2): 332-339.
KUO Y H, ZENG W T, CHEN J. Blind identification of primitive BCH codes parameters based on probability approximation[J]. Journal of Electronics & Information Technology, 2014, 36(2): 332-339.

[3] YU P D, LI J, PENG H. A least square method for parameter estimation of RSC sub-codes of Turbo codes[J]. IEEE Communications Letters, 2014, 18(4): 644-647.

[4] 刘骏, 李静, 于沛东. 一种 Turbo 码随机交织器的迭代估计方法[J]. 通信学报, 2015, 36(6): 2015140.
LIU J, LI J, YU P D. An iterative estimation method for random interleaver of Turbo codes[J]. Journal on Communications, 2015, 36(6): 2015140.

[5] 马钰, 张立民. 基于实时检测的扰码重建算法[J]. 电子与信息学报, 2016, 38(7): 1794-1799.
MA Y, ZHANG L M. Reconstruction of scrambler with real-time test[J]. Journal of Electronics & Information Technology, 2016, 38(7): 1794-1799.

[6] MACKAY D. Good error-correcting codes based on very sparse matrices[J]. IEEE Transactions on Information Theory, 1999, 45(2): 399-431.

[7] Wireless LAN medium access control (MAC) and physical layer (PHY) specifications amendment 5: enhancements for higher throughput[S]. IEEE Std. 802.11n-2009, 2009.

[8] XIA T, WU H C. Novel blind identification of LDPC codes using average LLR of syndrome a posteriori probability[J]. IEEE Transactions on Signal Processing, 2014, 62(3): 632-640.

[9] YU P D, PENG H, LI J. On blind recognition of channel codes within a candidate set[J]. IEEE Communications Letters, 2016, 20(4): 736-739.

[10] 包昕, 周磊硐, 何可, 等. LDPC 码稀疏校验矩阵的重建方法[J]. 电子科技大学学报, 2016, 45(2): 191-196.
BAO X, ZHOU L K, HE K, et al. A method of restructuring LDPC parity-check matrix[J]. Journal of University of Electronic Science and Technology of China, 2016, 45(2): 191-196.

[11] 包昕, 周磊硐, 何可, 等. 误码条件下的 LDPC 码盲识别算法[J]. 西安交通大学学报, 2015, 49(12): 53-58.
BAO X, ZHOU L K, HE K, et al. A recognition algorithm for LDPC codes of blind in a noisy environment[J]. Journal of Xi'an Jiaotong University, 2015, 49(12): 53-58.

[12] CLUZEAU M. Block code reconstruction using iterative decoding techniques[C]//Proceedings of IEEE International Symposium on Information Theory Seattle. USA, 2006: 2269-2273.

[13] CLUZEAU M, FINIASZ M. Recovering a code's length and synchronization from a noisy intercepted bit-stream[C]//IEEE International Symposium on Information Theory. Seoul, Korea, 2009: 2737-2741.

[14] CLUZEAU M, TILLICH J P. On the code reverse engineering problem[C]//IEEE International Symposium on Information Theory. Toronto, Canada, 2008: 634-638.

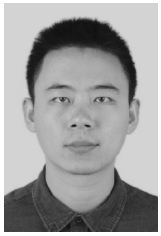
[15] CANTEAUT A, CHABAUD F. A new algorithm for finding minimum weight words in a linear codes: application to primitive narrow-sense BCH codes of length 511[J]. IEEE Transactions on Information Theory, 1998, 44(1): 367-378.

[16] CHEN L, XU J, DJURDJEVIC I, et al. Near-Shannon-limit quasi-cyclic low-density parity-check codes[J]. IEEE Transactions on Communications, 2004, 52(7): 1038-1042.

[17] 李开丁. 关于 n 个独立同分布的指数分布的最值问题的期望和方差[J]. 大学数学, 2005, 21(4): 125-127.
LI K D. About the mean and variance of the maximum and minimum of n independent exponential distributed random variables[J]. College Mathematics, 2005, 21(4): 125-127.

[18] CHABOT C. Recognition of a code in a noisy environment[C]//IEEE International Symposium on Information Theory. Nice, France, 2007: 2211-2215.

作者简介：



于沛东（1989-），男，湖南慈利人，解放军信息工程大学博士生，主要研究方向为信道编码及其识别分析。

彭华（1973-），男，江西萍乡人，解放军信息工程大学教授、博士生导师，主要研究方向为软件无线电、通信信号处理等。

巩克现（1976-），男，山东泰安人，解放军信息工程大学副教授、硕士生导师，主要研究方向为软件无线电、信道编码等。

陈泽亮（1992-），男，湖南岳阳人，解放军信息工程大学硕士生，主要研究方向为信道编码识别分析。